Supplementary Material for "SalChartQA: Question-driven Saliency on Information Visualisations"

Yao Wang
University of Stuttgart
Stuttgart, Germany
yao.wang@vis.uni-stuttgart.de

Weitian Wang
University of Stuttgart
Stuttgart, Germany
st176544@stud.uni-stuttgart.de

Abdullah Abdelhafez*
German University in Cairo
Cairo, Egypt
abdullahabdelhafez01@gmail.com

Mayar Elfares
University of Stuttgart
Stuttgart, Germany
mayar.elfares@vis.uni-stuttgart.de

Zhiming Hu†
University of Stuttgart
Stuttgart, Germany
zhiming.hu@vis.uni-stuttgart.de

Mihai Bâce*
KU Leuven
Leuven, Belgium
mihai.bace@kuleuven.be

Andreas Bulling
University of Stuttgart
Stuttgart, Germany
andreas.bulling@vis.uni-stuttgart.de

ACM Reference Format:
Yao Wang, Weitian Wang, Abdullah Abdelhafez, Mayar Elfares, Zhiming Hu, Mihai Bâce, and Andreas Bulling. 2024. Supplementary Material for "SalChartQA: Question-driven Saliency on Information Visualisations". In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3613904.3642942

This document contains implementation details of collecting SalChartQA dataset (Section 1), a screenshot of online study (Figure 1), participant accuracy distribution for all HIT responses (Figure 3), a screenshot of monitoring interface (Figure 2), the architecture of the decoder of VisSalFormer (Figure 4), and sample visualisation and questions for human ground truth and all baseline saliency maps (Figure 5, Figure 6, and Figure 7).

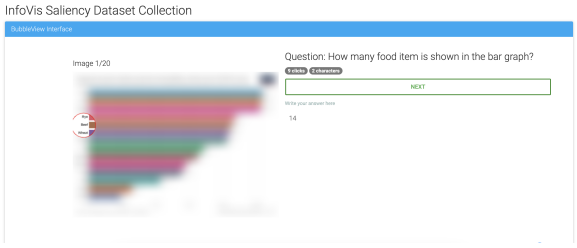

Figure 1: A screenshot of our online study. Blurry visualisations with a Gaussian kernel size of 40 px are shown to viewers alongside a question. Participants read the visualisation under the BubbleView setting, where mouse clicks to create a deblurred circle with 30 px radius.

1 DATASET COLLECTION

1.1 Web Development

A website was created using an open-source BubbleView codebase[1]. It incorporated several crucial components, including NeDB as the database management system to ensure efficient data storage and Express.js for constructing a RESTful API. Figure 2 illustrates the monitoring interface, a comprehensive platform for displaying and validating submissions from AMT. It offers clear data presentation and incorporates quality-of-life (QoL) features for improved usability. An essential function is visualizing participants' click coordinates, allowing for a precise understanding of their engagement patterns within the visualizations. Researchers can toggle

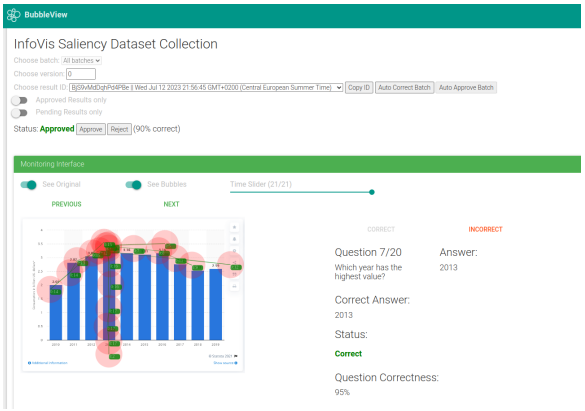

Figure 2: Monitoring interface.

*Work conducted at University of Stuttgart.
†Corresponding author
[1]https://github.com/namwkim/bubbleview

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/3613904.3642942

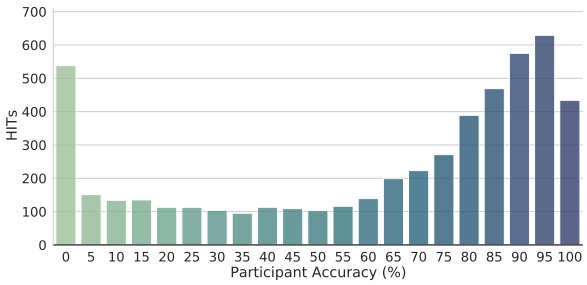

Figure 3: Participant accuracy over all 4,728 HIT responses.

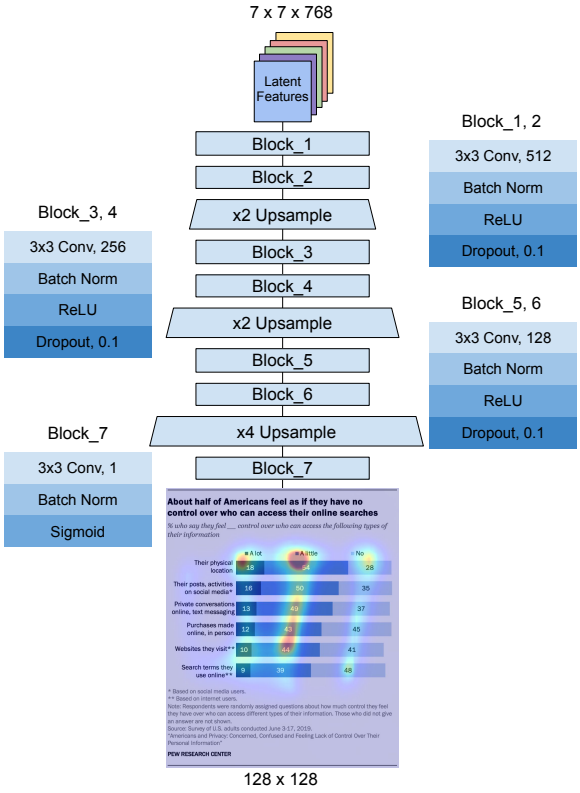

Figure 4: Architecture of the decoder of VisSalFormer.

between blurred and unblurred visualizations. Additionally, the interface streamlines manual data validation by providing an organized overview of submissions, expediting the identification and resolution of inconsistencies to ensure data quality and integrity.

1.2 Saliency Analysis of Discarded Participants

We conducted a within-in discarded participants' saliency analysis to check the data quality of discarded participants. For those participants who answered a question correctly, we randomly and evenly split them into two groups. The same split applied to those participants who answered a question wrongly. 547 questions were evaluated within rejected participants, with a minimum of two participants who answered a question correctly. A saliency map was generated for every group, and the similarity was computed between two saliency maps. The mean CC between correct-answer saliency maps was 0.411 ($\sigma = 0.266$), and 0.326 ($\sigma = 0.256$) between wrong-answer saliency maps. Considering the mean CC between correct-answer saliency maps was 0.760, and 0.549 for between wrong-answer saliency maps, the disagreement between discarded participants is higher than approved participants. It suggests the saliency maps of discarded participants encode much more noise and less question-related information than approved participants.

1.3 Question Keyword List

Question priority ranks from the top to the bottom in the following list. For example, if both "how many" and "maximum" appeared in a question, this question belongs to the calculation questions.

Keywords for comparison questions: than, compare, equal, more, difference, between, less, greater, higher, same, drop, gap, above, below.

Keywords for computing derived value questions: how many, calculate, add, substract, deduct, subtract,time, ratio, average, median, mode, total, sum, adding, summation, combined.

Keywords for data retrieval questions: what, percent, distribution, and value.

Keywords for finding extremum questions: maximum, minimum, peak, highest, lowest, largest, greatest, smallest, shortest, longest, tallest, biggest, most, least.

Keywords for filtering questions: which, represent, when, find.

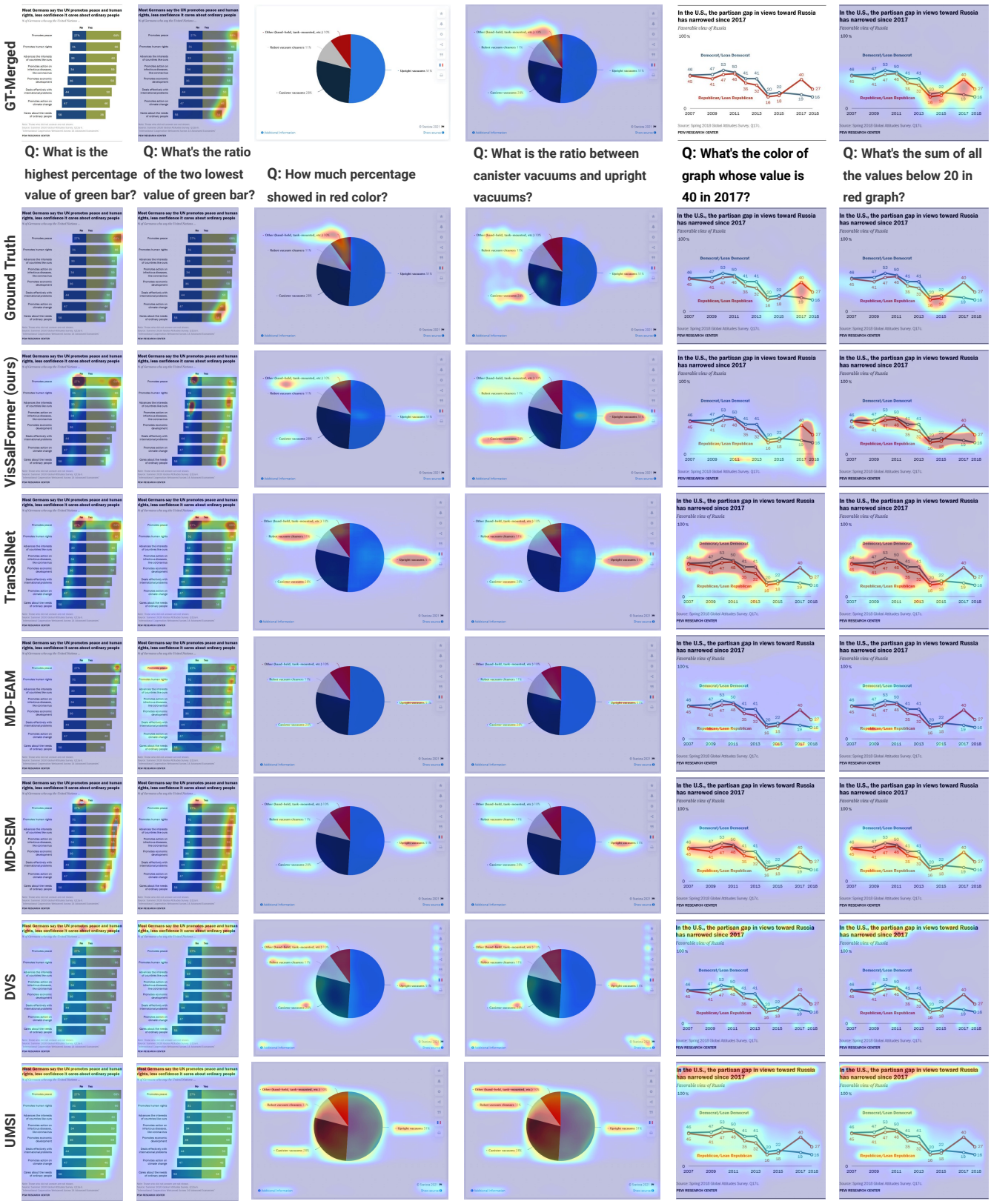

Figure 5: Sample visualisations and questions from the test set of SalChartQA with human ground truth saliency maps (top row). Predictions from our VisSalFormer and all baseline methods (TranSalNet, MD-EAM, MD-SEM, DVS, and UMSI) are shown below.

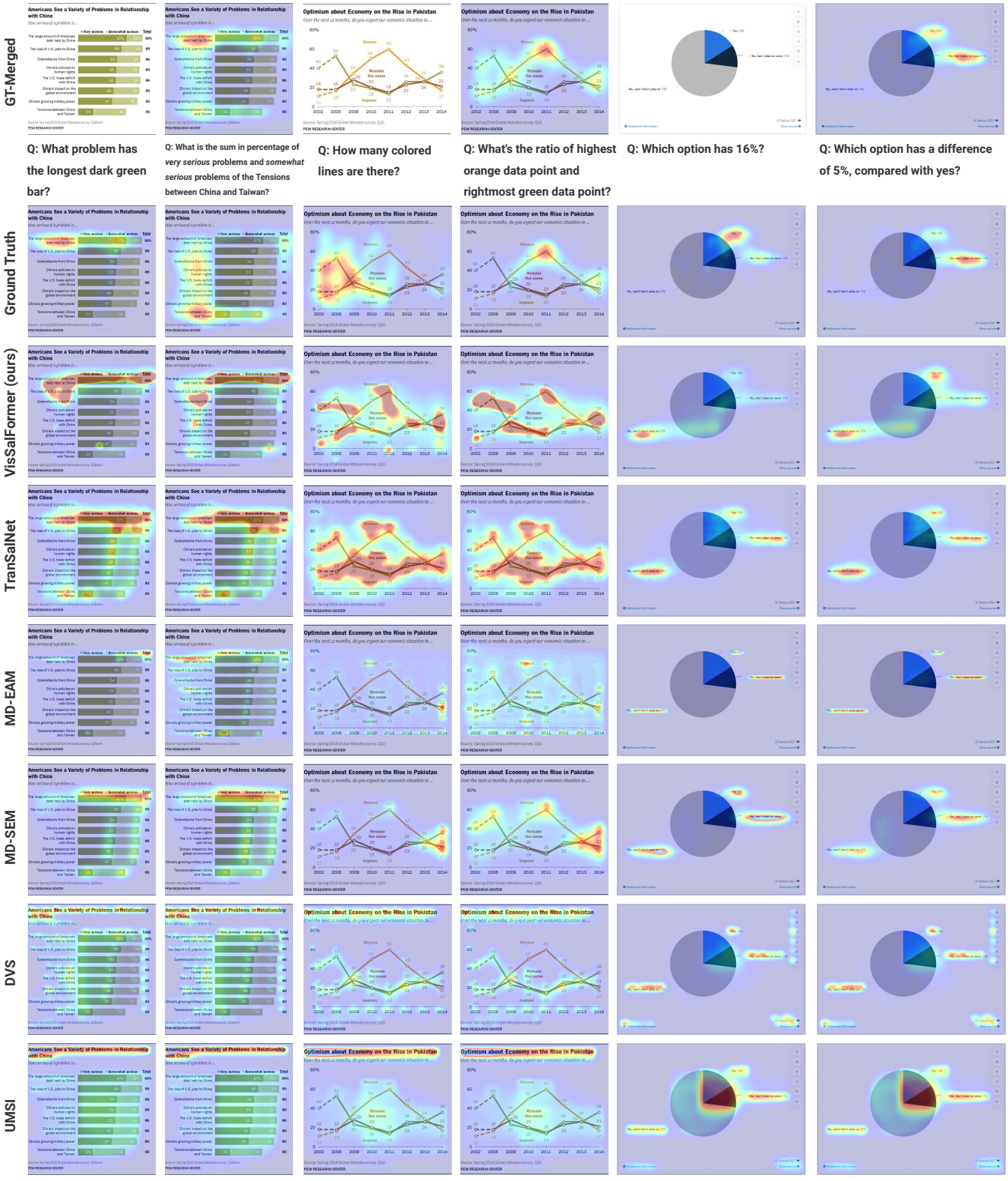

Figure 6: More visualisations and questions from the test set of SalChartQA with human ground truth saliency maps (top row). Predictions from our VisSalFormer and all baseline methods (TranSalNet, MD-EAM, MD-SEM, DVS, and UMSI) are shown below.

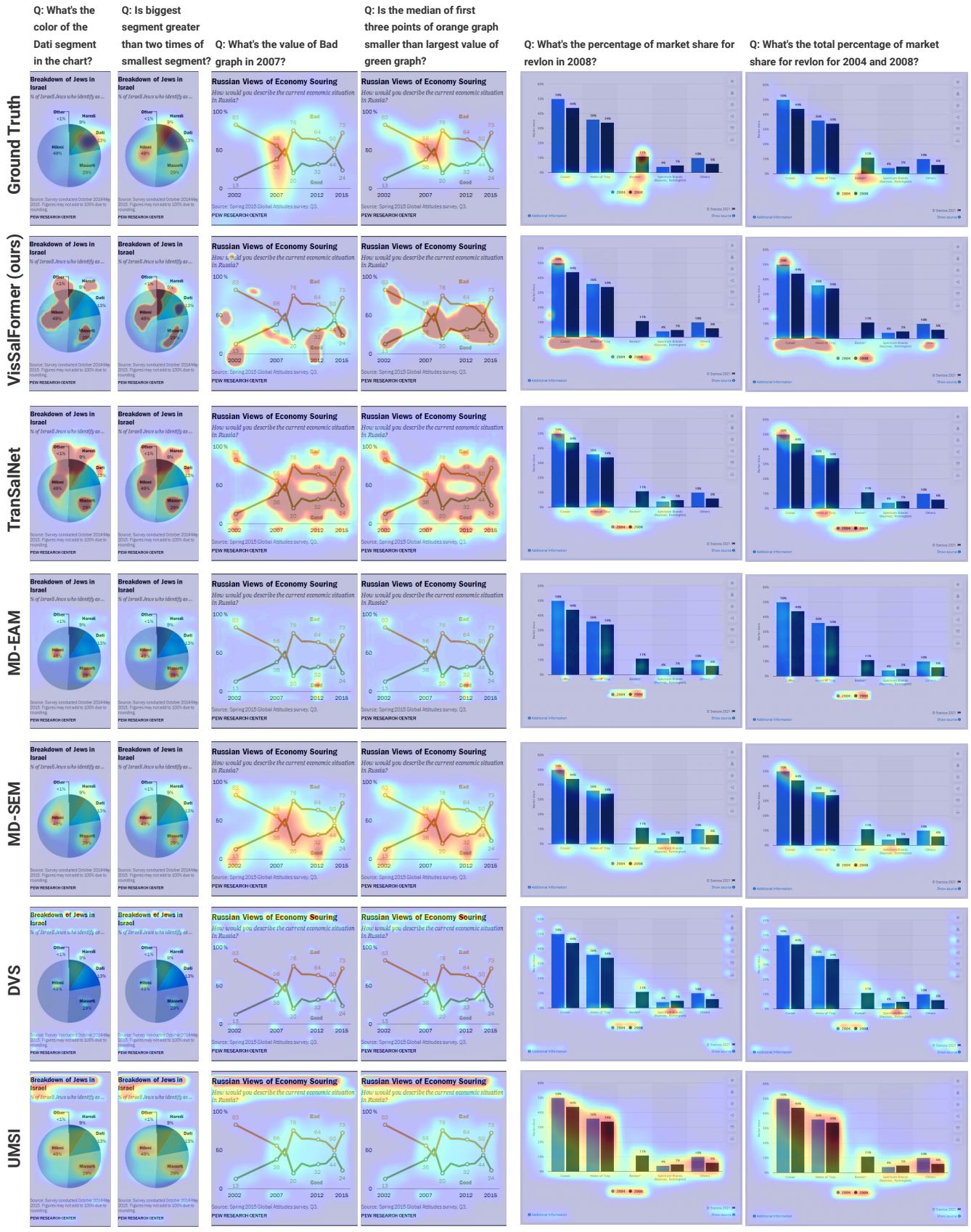

Figure 7: More visualisations and questions from the test set of SalChartQA with human ground truth saliency maps (top row). Predictions from our VisSalFormer and all baseline methods (TranSalNet, MD-EAM, MD-SEM, DVS, and UMSI) are shown below.